Notes on causal inference

Tri Nguyen
nguyetr9@oregonstate.edu

May 6, 2022

1 Why linear Gaussian is non-identifiable?

Theorem quoted in Theorem 4.2 [PJS17]

Theorem 1.1 (Linear Gaussian)**.** *Given that $Y = \alpha X + N_Y$, where $X \perp\!\!\!\perp N_Y$. Then there exists N_X and β such that $X = \beta Y + N_X$ such that $Y \perp\!\!\!\perp N_X$ if and only if both X and N_Y are normally distributed.*

Proof. We need to prove both directions.

Step 1. The $\Rightarrow$ direction: assume $X, Y_N \sim \mathcal{N}$.
So we need to find some β and N_X such that they makes Y and N_X to be independent.
Since $Y \sim \mathcal{N}$, it would be easier if N_X is also normal. In that case, we only need to design N_X so that

$$\mathrm{cov}(Y, N_X) = 0.$$

So the goal is to find N_X that satisfy:

- N_X is normal.
- $cov(Y, N_X) = 0$.

Let start from

$$N_X = X - \beta Y = X - \beta(\alpha X + N_Y) = (1 - \alpha\beta)X - \beta N_Y$$

haha, this is already normal little enforcement to ensure the whole thing is not 0. So the other thing should be

$$\begin{aligned}
& cov(N_X, Y) = 0 \\
\Leftrightarrow\quad & \mathrm{cov}((1-\alpha\beta)X - \beta N_Y, \alpha X + N_Y) = 0 \\
\Leftrightarrow\quad & (1-\alpha\beta)\alpha\mathrm{var}(X) + (1-\alpha\beta)\mathrm{cov}(X, N_Y) - \alpha\beta\mathrm{cov}(N_Y, X) - \beta\mathrm{var}(N_Y) = 0 \\
\Leftrightarrow\quad & (1-\alpha\beta)\alpha\mathrm{var}(X) - \beta\mathrm{var}(N_Y) = 0 \\
\Leftrightarrow\quad & \alpha\mathrm{var}(X) = \alpha^2\beta\mathrm{var}(X) + \beta\mathrm{var}(N_Y) \\
\Leftrightarrow\quad & \beta = \frac{\alpha\mathrm{var}(X)}{\alpha^2\mathrm{var}(X) + \mathrm{var}(N_Y)}
\end{aligned}$$

Step 2. The $\Leftarrow$ direction: assume the existence of N_X, β and ...It turns out the reverse direction is harder to prove. We need to invoke Darmois-Skitovich theorem.

Theorem 1.2 (Darmois-Skitovich)**.** *Let $X_1, \ldots, X_N$ be mutually independent. If there are exists nonzero coefficients α_i, β_i such that*

$$Y_1 = \sum_{i=1}^{N} \alpha_i X_i, \quad Y_2 = \sum_{i=1}^{N} \beta_i X_i$$

where $Y_1 \perp\!\!\!\perp Y_2$, then all X_i's are normally distributed.

Now we have two RV $X \perp\!\!\!\perp N_Y$, and there are some nonzero coefficients that make

$$\begin{cases} Y = \alpha X + N_Y \\ N_X = (1 - \alpha\beta)X - \beta N_Y \end{cases}$$

$Y \perp\!\!\!\perp N_X$. Hence by Darmois-Skitovich theorem, X, N_Y are normally distributed.
Of course, we need to handle some corner cases where some of the coefficients are 0. But since I'm lazy ... $\square$

Whenever two RVs are not independent, either one RV causes the other or either there is a common cause to both of them. In the former case, we want to determine the direction of the causality given joint pdf of 2 RVs. For example, if X causes Y, our basic assumption is $Y = f(X, N_Y)$ where the noise $N_Y \perp\!\!\!\perp X$. In general, if one tries to model $X = g(Y, N_X)$, the estimated noise N_X would satisfy $N_X \not\perp\!\!\!\perp Y$. So the difference in $N_Y \perp\!\!\!\perp X$ and $N_X \not\perp\!\!\!\perp Y$ are the key to determine direction of causality. However, in case that X, N_Y are independent and normal, and the model are linear, then the reverse model N_X, Y are also independent, as pointed out in Theorem 1.1. Hence, in that case, although the causal effect is unidirectional in reality, one cannot distinguish between 2 directions using only joint pdf of X, Y.

References

[PJS17] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms.* The MIT Press, 2017.